

Automatic Participant Evaluation in Computer Mediated Persistent Conversations

Field of the Invention

- [0001] The present invention relates to computer mediated persistent conversations such as Usenet newsgroups, email lists, and Web-based discussion boards and, in particular, to automatic evaluation of participants in such conversations.

Background and Summary of the Invention

- [0002] There are a wide variety of computer mediated persistent conversations, such as those that arise in various networked computer contexts, including Usenet (NNTP) newsgroups, World Wide Web (HTTP) message board or bulletin board sites, email lists, chat rooms, multi-user dungeons (MUDs), multi-user games and graphical virtual worlds, etc. These computer mediated persistent conversations are used singly and in concert to create an increasing number of social cyberspaces, groups, or communities that allow groups of people to gather on-line and interact with each other.
- [0003] While such computer mediated persistent conversations are becoming increasingly popular, there are a number of problems and difficulties facing the users of such media. The interaction context, or information, about the kind of space, group, and activity taking place in these media are frequently missing or ambiguous. This makes finding groups to participate in and people to interact with difficult. The history and structure of groups often become apparent only after extended participation. As a result, many of these media remain complex and confusingly tangled spaces in which it is difficult for users to navigate and participate.
- [0004] One example of such an interaction medium or environment is Usenet, which is a collection of messages, referred to as posts,

available through the Network News Transport Protocol (NNTP). The posts available through NNTP are organized into newsgroups that are named according to a hierarchical name convention. At the highest level there are numerous general categories, such as rec, comp, alt, sci, etc., that identify the general subject (i.e., recreation, computer, alternative, science, etc.) of a set of newsgroups. Additional terms are added hierarchically to uniquely define a newsgroup within a collection of others (e.g., comp.lang.perl.misc or rec.pets.misc, kids.pregnancy).

[0005] Public, online social spaces like Usenet newsgroups are frequently noisy and voluminous places. Larger groups regularly receive more than ten thousand messages a month; many of these messages being of limited value. Despite the many advantages groups of people gain when they interact through computer networks, enough poor-quality messages flood newsgroups and similar public discussions that finding the valuable content typically proves too difficult to be worthwhile.

[0006] Efforts to address this issue have primarily focused on methods for groups of people to share their opinions of the messages and authors contributing content to a social cyberspace. These techniques, used by Web sites like online auctioneer eBay.com and Linux news provider Slashdot.org, face the dual challenges of attracting a critical mass of people willing to contribute evaluations and ensuring the quality of those evaluations. Even when they meet these dual challenges, conventional network sites have not succeeded in using the information to reduce the noise and prevent the abuse that often prevail in online public spaces. The rampant misuse of these common spaces is one of the reasons many people who once used Usenet no longer do so.

[0007] Therefore, an aspect of the present invention is a determination that explicit participation or feedback of users may not be required to assess other users as to which authors and messages are valuable. This invention can automatically bring to online spaces or communities social context and interactional history that can help users evaluate other users.

[0008] In one implementation, the present invention provides information about the histories of the authors or users in a computer mediated persistent conversation, such as a newsgroup. The information can include how long a user has been active in the group, in what other newsgroups the user participates, to what other threads of conversation the user contributes, and which other users or participants the user most often engages in discussion. This information may be obtained as an aggregation of data that is obtained from the groups or communities automatically. The information about any selected user or group is then made available or otherwise provided to any other users so that they can decide whether the selected user or group is likely to provide helpful or useful information.

[0009] In an alternative implementation, particular message type indicators may be applied to messages when they are posted. The message type indicators may also be obtained as an aggregation of data obtained from the groups or communities automatically. The message type indicators may be made available, either alone or with any other information about any selected user or group, so that other users can decide whether the selected user or group is likely to provide helpful or useful information.

[0010] Additional objects and advantages of the present invention will be apparent from the detailed description of the preferred

embodiment thereof, which proceeds with reference to the accompanying drawings.

Brief Description of the Drawings

- [0011] Fig. 1 is a diagram illustrating an exemplary prior art computer display visualization of threaded posts or messages in a threaded information system such as a newsgroup.
- [0012] Fig. 2 is a flow diagram of a participant evaluation method for automatically evaluating participants in computer mediated persistent conversations.
- [0013] Fig. 3 is a block diagram of a participant evaluation system.
- [0014] Fig. 4 is a flow diagram of an alternative participant evaluation method for automatically evaluating participants in computer mediated persistent conversations.
- [0015] Figs. 5A-5E are illustrations of successive screens of a profile user interface that is rendered on a display screen of a user or client computer.
- [0016] Fig. 6 is a block diagram of a computer system that may be used to implement the present invention.

Detailed Description of Preferred Embodiments

- [0017] The present invention relates to computer mediated persistent conversations, such as those that arise in various networked computer contexts, including Usenet (NNTP) newsgroups, World Wide Web (HTTP) message board or bulletin board sites, email lists, chat rooms, etc. The following description refers to Usenet newsgroups, but is similarly applicable to other computer mediated persistent conversations.
- [0018] Usenet is a collection of messages, referred to as posts, available through the Network News Transport Protocol (NNTP). The posts available through NNTP are organized into newsgroups that are named according to a hierarchical name convention. At the highest level are 15 or more general categories, such as rec,

comp, alt, sci, etc., that identify the general subject (i.e., recreation, computer, alternative, science, etc.) of a set of newsgroups. Additional terms are added hierarchically to uniquely define a newsgroup within a collection of others (e.g., comp.lang.perl.misc or rec.pets.misc.kids.pregnancy). Many posts reference or reply to another post. As an organizational aid, posts that reference or reply to another post are nested or positioned directly under it to form a "thread." Threads are organizational trees with an initial post at a top level and branches to all the responsive posts.

[0019] Fig. 1 is a diagram illustrating an exemplary prior art computer display visualization 100 of threaded posts or messages 102 in a threaded information system such as a newsgroup or a Web-based message board. Every post 102 in visualization 100 is identified by text, such as the subject line or title for the post, author, and the date and time of the post. It will be appreciated, however, that each post or message in the newsgroup typically includes in its body text or other information (e.g., data files or hyperlinks) in addition to the subject line or title. Accordingly, it will be understood that posts 102 in visualization 100 are representations of underlying posted information (e.g., messages).

[0020] Posts 102 that reference other posts are arranged as threads 106. Posts that do not reference other posts may be referred to as top-level or initiating posts 102', which are a subset of posts 102. Top-level posts 102' initiate threads 106. Although a threaded information environment will commonly include many threads 106, at least some top-level posts 102' are typically not referenced by other posts 102.

[0021] The present invention is directed to helping users or readers of computer mediated persistent conversations (e.g., newsgroups) quickly find messages that they would actually like to read and avoid messages they would prefer not to read. For casual

newsgroup users, such a system would reduce the clutter of undesirable messages and point them to authors and messages that are more likely to be of value. More serious newsgroup users, especially those who primarily seek technical solutions, could benefit greatly from a system that helps them efficiently find what they are looking for amid the scores or even hundreds of daily posts.

[0022] Fig. 2 is a flow diagram of a participant evaluation method 200 for automatically evaluating participants in computer mediated persistent conversations, such as Usenet newsgroups. Fig. 3 is a block diagram of a participant evaluation system 300 that utilizes method 200. By the following description it will be appreciated that participant characterization method 200 may be computer-implemented in accordance with corresponding software instructions.

[0023] Process block 202 indicates that data (e.g., newsgroup posts and related information) are collected from a collection of conversations groups or newsgroups. The data may be collected by a conventional news server 302 (e.g., a conventional Usenet News (NNTP) server). In one exemplary implementation, news server 302 carries nearly 50,000 Usenet (NNTP) newsgroups, collecting all the messages in all the newsgroups. It will be appreciated, however, that news server 302 may carry fewer or more newsgroups and may collect less than entire messages in the newsgroups.

[0024] Process block 204 indicates that selected information is extracted from news server 302 and prepared for insertion into a database 304 (e.g., a SQL database). With reference to system 300, the selected information is extracted from news server 302 by a data collector 306. An example of data collector 306 is referred to as Netscan: A tool for measuring and mapping social

2007157E-020802

cyberspaces, and use of it is available from the assignee of the present invention at <http://netscan.research.microsoft.com>. The exemplary Netscan data collector 306 communicates with news server 302, obtaining selected information from the message headers for storage in database 304.

[0025] Process block 206 indicates that the selected information is stored in database 304.

[0026] Process block 208 indicates that one or more data scrubbing techniques are used to correct errors and anomalies in the raw data stored in database 304. Errors and anomalies that may be corrected include damage to thread structure data, localized variations in date and time stamps, and author identities. The data scrubbing techniques are directed to identifying such predefined errors and anomalies and corrects them by correcting thread data structure based on implicit or apparent thread relationships between messages, converting all date and time stamps to a common standard (e.g., Greenwich Mean Time), and discerning unique user identities based on user email addresses or combinations of email addresses and name aliases.

[0027] Process block 210 indicates that aggregation is applied to the data stored in database 304. As is known in the art, data aggregation (sometimes called “data scraping”) is a process in which information is gathered and expressed in a summary form. In one implementation, the aggregation applied to the data stored in database 304 generates a multidimensional data space with three primary dimensions: Time, Newsgroup, and Authors. For example, that multidimensional data space may include any or all of the following aggregations of message data:

Time: day, week, month, quarter, year, YTD, all resolutions

Newsgroup: leaf newsgroup, 1st, 2nd, and 3rd newsgroup level aggregations

Authors: 1st and 2nd level user domain and full email address

[0028] Examples of particular data aggregations that can be used to evaluate an author or newsgroup are listed below. It will be appreciated that any, or all, or other data aggregations that can be used to evaluate an author or newsgroup. The exemplary data aggregations are:

- The total number of newsgroups to which an author posted at least a single message.
- The dates the author was first and last seen in any newsgroup across the Usenet as well as in a selected or focal newsgroup.
- The number of days on which the author posted one or more messages to any newsgroup across the Usenet as well as in the focal newsgroup.
- The total number of messages the author posted to any newsgroup across the Usenet as well as the number posted to the focal newsgroup.
- The total number of messages the author posted to any newsgroup across the Usenet as well as the number posted to the focal newsgroup that were initial turns ("thread heads").
- The total number of messages the author posted to any newsgroup across the Usenet as well as the number posted to the focal newsgroup that were responses to other messages ("replies").
- The total number of other authors to whom the author replied in any newsgroup across the Usenet as well as the number replied to within the focal newsgroup ("reply targets").
- The total number of replies to messages posted by each author in any newsgroup across the Usenet as well as the number replied to within the focal newsgroup ("responses").
- The total number of other authors who responded to messages the author posted in any newsgroup across the Usenet as well

20071115-020000

as the number responded to within the focal newsgroup ("reply targets").

- The total number of threads to which the author posted at least a single message in any newsgroup across the Usenet as well as the number of threads in the focal newsgroup.
- The average number of messages each author posted in each thread they joined both in any newsgroup across the Usenet as well as within the focal newsgroup.
- The average number of generations ("depth") of turns and reply messages of each of the threads the author posted to in any newsgroup across the Usenet as well as in the focal newsgroup.
- The average number of sibling turns beneath each parent message ("breadth") in each of the threads the author posted to in any newsgroup across the Usenet as well as in the focal newsgroup.

[0029] Process block 212 indicates that the aggregation data are made available to users, such as in a communities portal network site or a profile that is provided with reference to a user or a newsgroup. For example, making the aggregation data available may include providing a network (e.g., World Wide Web) interface to the data via selected queries and reports that can provide including visualizations of threads, aggregations of newsgroups and data about authors. Also, a registration system may authenticate users and link them to a profile store. In addition, a series of customized queries can be generated to address a user's contexts and memberships. For example, a user may access different Usenet segments according to the user's different roles, such as a professional role relating to the user's work or profession, a parenting role relating to the user's interests as a

parent, or a gamer role relating to the user's interests in online and other computer-based games.

[0030] The data aggregations obtained by method 200 can provide participation metrics for evaluating or characterizing participants in computer mediated persistent conversations. The participation metrics relating to an author or group can be positive, or negative, indications of whether other the author or group is likely to provide useful information to other users. The participation metrics are objective characterizations from which other users can base subjective judgments about the quality or characteristics of information from a particular author or group.

[0031] Listed below are examples of participation metrics by which authors have been correlated positively with user impressions of the quality of information provided by an author:

- The number of days on which the author posted in the focal newsgroup and in all newsgroups
- The number of messages the author posted in the focal newsgroup and in all newsgroups
- The number of distinct threads in which the author participates in the focal newsgroup
- The depth and breadth of the threads in which the author participates
 - Depth of threads in the focal newsgroup
 - Breadth of threads in the focal newsgroup
 - Depth of threads in all newsgroups
 - Breadth of threads in all newsgroups
- The fraction of the author's messages which are replies, in the focal group and in all newsgroups
- The number of distinct people to whose messages the author replies in the focal newsgroup and in all newsgroups

- The number of replies which the author's messages garner in the focal group and in all newsgroups
- The number of distinct people who reply to the author's messages in the focal group and in all newsgroups

[0032] As an example, authors with relatively high values for the following participation metrics may commonly be deemed by others to be high quality participants who contribute meaningfully to a group:

- The fraction of the author's messages which are replies in the focal newsgroup and across all newsgroups
- The number of replies, reply targets, responses, and response targets across all newsgroups
- The number of days on which the author posted in the focal newsgroup and across all newsgroups

An author with high values for these participation metrics may be interpreted as an author who participates actively and regularly in a variety of in-depth conversations, in which he or she responds to other participants but does not overwhelm the discussion, or, participate in too many different newsgroups.

[0033] As another example, authors with relatively high values for the following participation metrics may commonly be considered negatively or unfavorably by others: the total number of newsgroups in which an author has posted and the average number of messages the author posted in the threads in which he or she participated. A high value for the latter participation metric can represent a person who dominates the conversations that they join.

[0034] Fig. 4 is a flow diagram of an alternative participant evaluation method 400 for automatically evaluating participants in computer mediated persistent conversations, such as Usenet

[illegible][illegible][illegible]

- [0037]** An aspect of the Closed message type is that it acknowledges the author who posts the resolving Answer, based upon the subjective judgment of the author who posed the original Question. As described below, this acknowledgement can function as a participation incentive for authors to provide high-quality, substantive contributions.
- [0038]** Process block 404 indicates that data (e.g., newsgroup posts) are collected from a collection of newsgroups, the data including the message type identifiers. For example, the data may be collected by a conventional news server 302 (e.g., a conventional Usenet News (NNTP) server) that is modified to also collect the message type identifiers.
- [0039]** Process block 406 indicates that selected information is extracted from news server 302 and prepared for insertion into database 304, the selected information including the message type identifiers.
- [0040]** Process block 408 indicates that the message type identifier information is stored in database 304.
- [0041]** Process block 410 indicates that one or more data scrubbing techniques are used to correct errors and anomalies in the raw data stored in database 304.
- [0042]** Process block 412 indicates that aggregation is applied to the message type identifier information stored in database 304, the aggregation including correlating the message type identifier information with corresponding authors.
- [0043]** Process block 414 indicates that the aggregation data are made available to users, such as in a communities portal network site or a profile that is provided with reference to a user or a newsgroup. It will be appreciated that the aggregation data may also be made available to users as a network service that is accessible by software on behalf of the user, rather than being

accessed by the user manually, such as in program-to-program integration over a network like the Internet.

[0044] An aspect of alternative participant evaluation method 400 is that it can function as an incentive for providing high-quality, substantive contributions to computer mediated persistent conversations. The combination of message type identifiers with acknowledgement of resolving Answers allows authors to establish a positive reputation for resolving Questions based upon feedback from other participants in a group. Such positive reputations can allow authors to distinguish themselves as experts or as being helpful and thereby encourage other users to seek the professional advise or assistance of these authors.

[0045] Figs. 5A-5E are illustrations of successive screens of a profile user interface 500 that is rendered on a display screen of a user or client computer for providing a user with aggregation data relating to a user or group. It will be appreciated that profile user interface 500 illustrates the aggregation data rendered in one exemplary implementation of the operation of process blocks 212 and 412. It will be appreciated that profile user interface 500 could include any or all of the illustrated data aggregations, as well as any of the others referred to or described herein.

[0046] Profile user interface 500 includes a user identification segment 502 that typically includes at least a user screen name, but may also include any or all of the user's actual name, email address, and any other contact or identifying information. Initial and most recent date fields 504 and 506 indicate when the user first joined a group or "community" of groups and when the user last visited the group or community, respectively.

[0047] A message type summary segment 508 may indicate counts of specified message types that the user has contributed to the group or community. In the illustrated implementation, message

type summary segment 508 may include counts of Questions message types posted, Answer message types posted, Answer message types designated as Closing a question, Review message types, and Event message types. A "last updated" field may also be included to indicate when message type summary segment 508 was last updated.

[0048] A questions posted segment 510 may be included to list questions or other initiating messages posted by the user in the group or community, together with any replies that have been posted. A group summary segment 512 may also be included for each group or community that the user specifies or participates in. The group summary segment 512 provides background on the user's interests and may include various characterizations of the group or community, including summary characterizations of frequent contributors, listings of active conversations or threads, an identification of most recent messages of one or more message types.

[0049] Fig. 6 illustrates an operating environment for an embodiment of the present invention as a computer system 620 with a computer 622 that comprises at least one high speed processing unit (CPU) 624 in conjunction with a memory system 626, an input device 628, and an output device 630. These elements are interconnected by at least one bus structure 632.

[0050] The illustrated CPU 624 is of familiar design and includes an ALU 634 for performing computations, a collection of registers 636 for temporary storage of data and instructions, and a control unit 638 for controlling operation of the system 620. The CPU 624 may be a processor having any of a variety of architectures including Alpha from Digital, MIPS from MIPS Technology, NEC, IDT, Siemens, and others, x86 from Intel and others, including Cyrix, AMD, and Nexgen, and the PowerPC from IBM and Motorola.

[0051] The memory system 626 generally includes high-speed main memory 640 in the form of a medium such as random access memory (RAM) and read only memory (ROM) semiconductor devices, and secondary storage 642 in the form of long term storage mediums such as floppy disks, hard disks, tape, CD-ROM, flash memory, etc. and other devices that store data using electrical, magnetic, optical or other recording media. The main memory 640 also can include video display memory for displaying images through a display device. Those skilled in the art will recognize that the memory 626 can comprise a variety of alternative components having a variety of storage capacities.

[0052] The input and output devices 628 and 630 also are familiar. The input device 628 can comprise a keyboard, a mouse, a physical transducer (e.g., a microphone), etc. The output device 630 can comprise a display, a printer, a transducer (e.g., a speaker), etc. Some devices, such as a network interface or a modem, can be used as input and/or output devices.

[0053] As is familiar to those skilled in the art, the computer system 620 further includes an operating system and at least one application program. The operating system is the set of software which controls the computer system's operation and the allocation of resources. The application program is the set of software that performs a task desired by the user, using computer resources made available through the operating system. Both are resident in the illustrated memory system 626.

[0054] In accordance with the practices of persons skilled in the art of computer programming, the present invention is described below with reference to acts and symbolic representations of operations that are performed by computer system 620, unless indicated otherwise. Such acts and operations are sometimes referred to as being computer-executed and may be associated

with the operating system or the application program as appropriate. It will be appreciated that the acts and symbolically represented operations include the manipulation by the CPU 624 of electrical signals representing data bits which causes a resulting transformation or reduction of the electrical signal representation, and the maintenance of data bits at memory locations in memory system 626 to thereby reconfigure or otherwise alter the computer system's operation, as well as other processing of signals. The memory locations where data bits are maintained are physical locations that have particular electrical, magnetic, or optical properties corresponding to the data bits.

[0055] Having described and illustrated the principles of our invention with reference to an illustrated embodiment, it will be recognized that the illustrated embodiment can be modified in arrangement and detail without departing from such principles. It should be understood that the programs, processes, or methods described herein are not related or limited to any particular type of computer apparatus, unless indicated otherwise. Various types of general purpose or specialized computer apparatus may be used with or perform operations in accordance with the teachings described herein. Elements of the illustrated embodiment shown in software may be implemented in hardware and vice versa.

[0056] In view of the many possible embodiments to which the principles of our invention may be applied, it should be recognized that the detailed embodiments are illustrative only and should not be taken as limiting the scope of our invention. Rather, we claim as our invention all such embodiments as may come within the scope and spirit of the following claims and equivalents thereto.